

# Breeding for Bio-ethanol Production in *Lolium perenne* L.: Association of Allelic Variation with High Water-Soluble Carbohydrate Content

Kerrie Farrar · David N. Bryant · Lesley Turner ·  
Joe A. Gallagher · Ann Thomas · Markku Farrell ·  
Mervyn O. Humphreys · Iain S. Donnison

Published online: 19 October 2011

© The Author(s) 2011. This article is published with open access at Springerlink.com

**Abstract** Increasing the extractable sugar yield from perennial crops is one strategy to generate renewable fuels such as bio-ethanol. *Lolium perenne* L. (perennial ryegrass) can contain significant (>30% dry matter) water-soluble sugars in the form of polymeric fructan which is readily extracted, broken down and fermented to bio-ethanol. A population of *L. perenne* generated from four parents which differed in water-soluble carbohydrate (WSC) content was subjected to multiple rounds of selection and recombination on the basis of early spring WSC content to produce a high WSC, and a low WSC population. A control population was generated by selecting the same number of plants at random. The alleles present at six candidate gene loci were analysed before and after selection and correlated to WSC content. Significant differences in the allele frequency of *L. perenne soluble-acid invertase1:4* were observed between the three populations with one haplotype significantly associated with the high WSC C2<sup>S+</sup> population (after three rounds of selection and two rounds of recombination). Moreover, WSC content was also associated with biomass accumulation. Thus, in addition to a 2.84-fold increase in WSC yield, the C2<sup>S+</sup> population also had 1.48-fold more biomass per plant, resulting in 3.9-fold higher WSC yield per plant than the control population.

**Keywords** Single-nucleotide polymorphisms (SNPs) · Marker-assisted selection (MAS) · Phenomics · Single-strand conformational polymorphism (SSCP) · *Invertase1:4* (*LpsaINV1:4*)

## Abbreviations

<i>Lp6G-FFT</i>	6-Glucose–fructose fructosyl transferase
BAC	Bacterial artificial chromosome
<i>LpcwINV1</i>	Cell wall invertase 1
<i>LpcwINV2</i>	Cell wall invertase 2
<i>LpC4H</i>	Cinnamate 4-hydroxylase
DMY	Dry matter yield
QTL	Quantitative trait loci
SNPs	Single-nucleotide polymorphisms
SSCP	Single-strand conformational polymorphism
<i>LpsaINV1:4</i>	Soluble-acid invertase 1:4
<i>LpsaINV5</i>	Soluble-acid invertase 5
WSC	Water-soluble carbohydrate

## Introduction

In order to decrease the use of fossil fuels, alternative technologies are being developed and deployed. In temperate climates, the use of perennial forage grasses with high sugar yield, and high biomass potential, for fermentation into bio-ethanol, is one approach to meet demands for renewable liquid fuel. *Lolium perenne* L. (perennial ryegrass), a major perennial forage grass in Europe and other temperate regions, is an ideal dual use or bioenergy crop having previously been bred for high water-soluble carbohydrate (WSC) content as well as high digestibility [1]. Changes in policy drivers, such as the Common Agricultural Policy reform in Europe, have reduced livestock stocking numbers and so there is now an

**Electronic supplementary material** The online version of this article (doi:10.1007/s12155-011-9156-0) contains supplementary material, which is available to authorized users.

K. Farrar · D. N. Bryant · L. Turner · J. A. Gallagher (✉) ·  
A. Thomas · M. Farrell · M. O. Humphreys · I. S. Donnison  
Institute of Biological, Environmental and Rural Sciences,  
Aberystwyth University, Gogerddan,  
Aberystwyth, Ceredigion SY23 3EB, UK  
e-mail: jbg@aber.ac.uk

K. Farrar  
e-mail: kkf@aber.ac.uk

opportunity for farmers to manage their grasslands to deliver both livestock and energy outputs, for example by taking an early spring cut for ethanol production prior to growing the grass for pasture, hay or silage.

*L. perenne* is a perennial temperate grass species which can accumulate WSC levels equivalent to, and higher (>30% dry matter (DM)) than those in sugarcane. These sugars are readily extracted by juicing the fresh biomass and easily broken down and fermented with little energy input required. The concentration of WSC in *L. perenne* varies during the year, it is highest in summer, but differences between high and low lines can be identified in spring.

Previous studies have identified a number of QTL for WSC in *L. perenne* [2, 3]. These have been primarily attributed to elevated concentrations of fructan, a water-soluble non-structural carbohydrate that is synthesised directly from sucrose and readily accumulated in *Lolium* spp. as an alternative polysaccharide to starch [2, 4, 5]. Association of candidate genes with QTL and subsequent marker development would enable molecular discrimination between plants harbouring alleles conferring differential WSC accumulation and hence be of value in breeding programmes to select high WSC plants.

Glycoside hydrolases (EC 3.2.1.-), comprise a ubiquitous and expanding super-family of sugar-releasing enzymes, which have a broad range of structures and substrate specificities. The glycoside hydrolase group of enzymes includes invertases, fructosyltransferases and fructan hydrolases which are all highly homologous. Fructosyltransferases are involved in the synthesis of fructans while fructan hydrolases are involved in their remobilisation. The fructans which accumulate in *L. perenne* are produced by the action of a number of fructosyltransferases, a sucrose/sucrose 1-fructosyltransferase (EC 2.4.1.99), a fructan/fructan 1-fructosyltransferase (EC 2.4.1.100), a fructan/fructan 6G-fructosyltransferase (EC 2.4.1.243) and a sucrose/fructan or fructan/fructan 6-fructosyltransferase (EC 2.4.1.10) [4]. Invertases (EC 3.2.1.26) hydrolyse sucrose to glucose and fructose and comprise both acid and alkaline isoforms depending on their pH for optimal activity, with acid invertases further classified as either soluble or cell wall invertases.

Glycoside hydrolases of microbial and plant origin are also of growing interest in biofuel production due to their ability to catalyze the release of fermentable sugars from both plant polysaccharide reserves and lignocellulosic material [5].

In this study, six gene loci (five glycoside hydrolases and a lignin biosynthesis gene) were selected and tested for the potential to predict WSC yields either in isolation or in combination. The results are discussed in relation to breeding *L. perenne* for bio-ethanol production.

## Methods

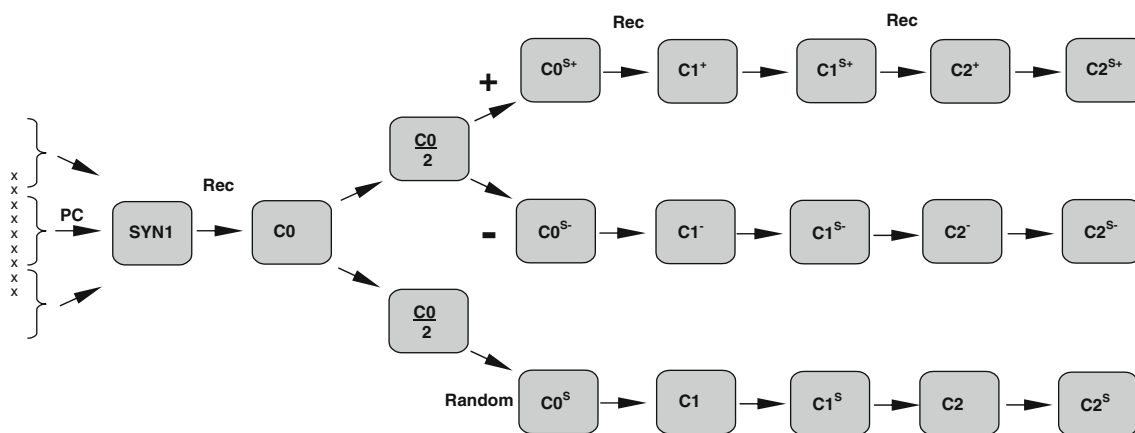
A synthetic population, generated from four parental lines comprising both high and low WSC genotypes, was subjected to successive rounds of selection on the basis of early spring WSC content followed by recombination, and the candidate gene loci analysed in order to assess shifts in allele frequency and thereby associate phenotype with genotype.

### Synthetic Populations and WSC Selection

A synthetic C0 generation was created by pair-crossing *L. perenne* parents LTS01, LTS05, LTS09 and LTS18 [6] followed by a round of recombination. LTS09 and LTS18 are from the same mapping family and have moderate to high WSC phenotypes. LTS01 and LTS05 are mapping family parents from The Netherlands and Denmark, respectively, and have low WSC phenotypes.

The C0 population ( $n=600$ ) was randomly divided in two. One half ( $n=300$ ) was subjected to divergent phenotypic selection. Three rounds of selection for high WSC and low WSC content were carried out in early March in years 2005, 2006 and 2007. Synthetic crosses (rounds of recombination) were carried out during the summers of 2005 and 2006. The three rounds of selection and two rounds of crossing produced a high WSC ( $C2^{S+}$ ) and a low WSC ( $C2^{S-}$ ) population. A selection intensity of 10% was used with 30 plants taken for each round of recombination from the population of 300. The second half of the population comprised the control population ( $n=300$ ) which underwent three rounds of random selection and two of recombination at the same intensity and in the same environment as the selected groups at each generation (Fig. 1).

Plants were grown and scored for WSC and dry matter yield (DMY) in the same environment. Poly-crosses for recombination were carried out in pollen-proof glasshouses. Seed was sown in August immediately following harvest, and phenotypic measurements for the next round of selection carried out the following spring. The resulting plants were maintained in 9-cm pots of Humax John Innes number 3 with wetting agent in a frost-free, unlit glasshouse throughout the year. The WSC and DMY data presented in this study are from plants of the same age collected at the same time of year. All samples within each year were harvested on the same morning during the first week of March from plants grown in an unlit glasshouse. C0, C1 and C2 data were collected on consecutive years (2005, 2006 and 2007, respectively), so natural environmental variation would be expected between years. Thus, the different populations of each generation can be directly compared, but absolute values between different generations cannot.



**Fig. 1** Generation of synthetic populations of *L. perenne* undergoing divergent WSC selection for the identification of candidate gene single nucleotide polymorphism distributions. (*PC* pairwise crosses, *SYN1* synthetic population 1, *C0* starting population ( $n=600$ ), *C1* generation

1 ( $n=300$ ), *C2* generation 2 ( $n=300$ ), *Rec* recombination, *S* selection ( $n=30$ ); *plus sign* high WSC; *negative sign* low WSC, *Random* not selected for on the basis of WSC)

### Biomass Harvest, Carbohydrate Extraction and Analysis

WSC was analysed in total herbage from a cut at approximately 4-cm stubble height. Total above-ground biomass harvested for this study was, therefore, predominantly leaf and also some sheath/tiller base material. The material was oven-dried to a constant mass at 80°C and WSC was extracted following the method of Turner et al. [7] but with the addition of 1.6 mg mL<sup>-1</sup> tartrazine to the extraction medium as an internal standard in order to account for variation in pipetting volumes during sample dilution and measurement. Total WSC of the extract was analysed directly with the anthrone colour reaction using a titre plate assay modified from Laurentin and Edwards [8], against 0–10 µg well<sup>-1</sup> fructose standards. The absorbance at 405 nm of a 40-µL sample (diluted at a rate of 1 µL in 160 µL) was recorded before the anthrone reactions, and used to correct plate sample volumes. Titre plates were first sealed in cling film and frozen at -20°C. Anthrone reagent (100 µL 2 mg mL<sup>-1</sup> anthrone in cold concentrated H<sub>2</sub>SO<sub>4</sub>) was added to unwrapped plates on an ice bath. They were sealed with acetate film, vortex mixed and the colour reaction carried out on a 95°C water bath. Plates were cooled, dried and if necessary centrifuged at 2000×g at room temperature for 1 min to remove any small bubbles on the base of the wells which would interfere with absorbance measurement. Data were expressed as mg g<sup>-1</sup> DM.

### Identification of Sequence Polymorphisms, and Single-Strand Conformational Polymorphism Analysis

A bacterial artificial chromosome (BAC) library with five times coverage of the *L. perenne* genome [9] was screened by using the polymerase chain reaction (PCR), with

forward and reverse oligonucleotide primers designed to candidate gene sequences. BACs containing genes of interest were partially sequenced. Approximately 1 kb of sequence, including a proportion of the promoter and/or intron where possible, was amplified from the five parental genotypes in order to identify polymorphisms. The primer sequences used are listed in Table 1. Following alignment, regions of approximately 250 bp containing multiple SNPs, insertions or deletions identified in the *L. perenne* parents were amplified for haplotype analysis by single-strand conformational polymorphism (SSCP) analysis using the primers listed in Table 1.

DNA was extracted from individual plants using the QIAGEN DNEasy Plant Mini Kit. SSCP analysis was performed by IDna Genetics Ltd, Norwich, NR4 7UH, UK. Briefly, 5' fluorescent-labelled gene-specific primers (Invitrogen, UK) were used to produce amplicons by PCR. Forward primers were labelled with 6FAM and reverse primers were labelled with VIC. PCR was conducted on the five C0 parental genotypes LTS01, LTS05, LTS09, LTS17 and LTS18 and on the 90 C2 genotypes (30 of each population: high WSC, low WSC and random). Amplicons were separated on an Applied Biosystems ABI3730 genetic analyzer equipped with a modified CAP polymer. Migration sizes were recorded and compared with the haplotype sequence.

### Data Analysis

Phenotype statistical analysis was conducted in Sigma plot 11.0. One-way ANOVA was used to determine within-year significant differences in WSC content and yield between recombinants and populations selected for high and low WSC levels along with the inclusion of the random selected WSC population at the completion of this study. Haplotype

**Table 1** List of PCR primer pairs used in this study

Gene locus	Amplicon	Forward primer	Reverse primer
<i>Cell-wall invertase 1 (LpcwINV1)</i>	gene locus	5' ccc agg aag ata tgg ctt ga	5' gcc gta tcc atc ctt gaa ga
<i>Soluble-acid invertase 1:4 (LpsaINV1:4)</i>	gene locus	5' cat cca aga cct tct atg ac	5' ttc taa aca act tgg act gac g
<i>Cell-wall invertase 2 (LpcwINV2)</i> and <i>soluble-acid invertase 5 (LpsaINV5)</i>	gene locus	5' ceg ggg ttc cat ttc cag cc	5' ten ayg cac tcc cac at
<i>6-glucose-fructose fructosyl transferase (Lp6G-FFT)</i>	gene locus	5' cct agt gtt taa tta ttg gac	5' cgg caa caa aga ttg gag c
<i>Cinnamate 4-hydroxylase (LpC4H)</i>	gene locus	5' ggc aac ttg ctg cag gtg	5' gat gtt ctc gac gat gta g
<i>Cell-wall invertase 1 (LpcwINV1)</i>	SSCP product	5' gaa tgc acg gtg ctc tc	5' gtc cat gtc gcc atc gat g
<i>Soluble-acid invertase 1:4 (LpsaINV1:4)</i>	SSCP product	5' gga ctt ctc cta caa tea ggc c	5' ttc taa aca act tgg act gac g
<i>Cell-wall invertase 2 (LpcwINV2)</i>	SSCP product	5' gta ata cat cgg ggg tgc ac	5' tgt tca gct ctt cgt cca aa
<i>Soluble-acid invertase 5 (LpsaINV5)</i>	SSCP product	5' ggg gtt gtg gct ggt aga	5' gca gac tca aat tct gct caa
<i>6-glucose-fructose fructosyl transferase (Lp6G-FFT)</i>	SSCP product	5' ctt ctt ggc tga cga cat g	5' cat gaa cca aca cct tct g
<i>Cinnamate 4-hydroxylase (LpC4H)</i>	SSCP product	5' ggc aac ttg ctg cag gtg	5' gag aca acg tgg tcc atc tgc

analysis was performed using population genetics software Arlequin 3.1 [10]. F-statistics were computed using 10,000 permutations for significance with 1000 permutations for the Mantel test. The exact test of population differentiation was computed with 500,000 Markov chain steps, 3000 dememorisation steps and a significance threshold of 0.01. Markers found to significantly differ between populations were scored for the haplotype number in each of the three populations, C2<sup>S+</sup>, C2<sup>S-</sup> and C2<sup>S</sup> random, in addition to the five initial *L. perenne* parental lines. Statistical analysis was performed using the Chi-square test with 2 degrees of freedom [11].

## Results

A synthetic population was generated from high and low WSC parents and submitted to three rounds of selection and two rounds of recombination on the basis of WSC content (Fig. 1) in order to identify loci associated with high WSC yields in *L. perenne*.

### Effect of Selection on WSC Content (mg g Dry Matter<sup>-1</sup>)

The diversity in WSC content within the starting population was such that the top 10% of plants had a WSC content 2.0-fold higher than the mean and 5.4-fold higher than the lowest 10% of plants (Table 2). Recombination within the two populations resulted in C1 populations ( $n=300$ ) in which the WSC values for the C1<sup>+</sup> population were 1.1-fold higher than the C1<sup>-</sup> population. Selection of the top 10% of plants from the C1<sup>+</sup> population resulted in a mean WSC yield that is 5.5-fold higher than the lowest 10% of plants of the C1<sup>-</sup> population. Following the final round of recombination and

selection, the C2<sup>S+</sup> population had a mean WSC content of 250.2±5.8 mg g DM<sup>-1</sup>, 12.5-fold higher than the C2<sup>S-</sup> population (Table 2). In all cases, the genetic diversity present within the selected populations resulted in an average WSC content closer to that of the C0 population following recombination in the subsequent generation, thus emphasising the need for recurrent selection in a breeding programme directed towards high WSC in order to eliminate alleles encoding low WSC phenotypes.

When compared with the C2<sup>S</sup> random population the WSC content of both the C2<sup>S+</sup> and the C2<sup>S-</sup> had undergone significant divergence ( $p<0.05$ ) from the WSC content of the C2<sup>S</sup> random population. In the three final populations C2<sup>S+</sup> had a WSC content 2.8-fold higher than the random population and the C2<sup>S-</sup> population had a WSC content 0.2 times that of the random population (Table 2).

### Effect of WSC Selection on Early Spring Dry Matter Yield

A secondary effect of selecting high WSC plants was the observation of increased DMY per plant compared with the controls. In addition to the expected differences in WSC content, the high and low WSC selected populations for C1 and C2 showed significant differences ( $p<0.05$ ) for DMY per plant, indicating that selection for WSC also resulted in increased DMY. Over the three generations, C0 to C2, the DMY of the high and low WSC populations had diverged such that the C2<sup>S+</sup> plants selected for high WSC demonstrated 2.3-fold higher DMY per plant than the C2<sup>S-</sup> population. In addition, the total DMY in C2<sup>S+</sup> high WSC selected plants was 1.5-fold higher than the C2<sup>S</sup> random population that had not undergone WSC directed selection. The DMY of the C2<sup>S-</sup> low WSC plants was 0.6 times that of the random population (Table 2).

**Table 2** WSC content, biomass and WSC yield, ± standard deviation, of the high (C+), low (C-) and random populations (n=300) through two rounds of WSC directed selection, and subsequent recombination (S=selected (n=30), ND=not determined)

	C0	C0 <sup>S</sup>	C1	C1 <sup>S</sup>	C2	C2 <sup>S</sup>	C2 <sup>S</sup> <i>LpsaINV1:4</i> haplotypes		
							251	252	251/252
Starting population WSC mg DM <sup>-1</sup>	100.5±2.87	–	–	–	–	–	–	–	–
C+ WSC mg g DM <sup>-1</sup>	–	*196.1±6.8	*148.0±3.0	*248.0±7.3	*149.2±2.6	*250.2±5.8	6	14	9
C- WSC mg g DM <sup>-1</sup>	–	*36.2±1.1	*129.0±2.8	*45.2±1.8	*66.0±1.3	*20.1±1.2	21	0	9
Random WSC mg g DM <sup>-1</sup>	–	ND	ND	ND	ND	*88.1±7.8	16	1	13
Starting population Biomass g	1.12±0.03	–	–	–	–	–	–	–	–
C+ Biomass g	–	**1.34±0.07	*1.20±0.03	*1.49±0.08	*1.55±0.02	*1.73±0.08	–	–	–
C- Biomass g	–	**0.7±0.05	*1.04±0.03	*0.48±0.05	*0.96±0.02	*0.74±0.05	–	–	–
random Biomass g	–	ND	ND	ND	ND	*1.17±0.06	–	–	–
Starting population WSC yield g plant <sup>-1</sup>	0.121±0.01	–	–	–	–	–	–	–	–
C+ WSC yield g plant <sup>-1</sup>	–	0.262±0.018	0.166±0.007	0.366±0.021	0.226±0.007	0.43±0.019	–	–	–
C- WSC yield g plant <sup>-1</sup>	–	0.026±0.002	0.128±0.006	0.023±0.002	0.064±0.003	0.015±0.001	–	–	–
random WSC yield g plant <sup>-1</sup>	–	ND	ND	ND	ND	0.109±0.001	–	–	–
C+/C ratio	–	10.1	1.3	15.9	10.8	28.7	–	–	–
C+/random ratio	–	ND	ND	ND	ND	3.9	–	–	–
C-/random ratio	–	ND	ND	ND	ND	0.14	–	–	–

One way ANOVA \*\* = significant difference between populations within each year ( $p < 0.001$ ); one way ANOVA on ranks, pair wise comparison \* = significant difference between populations within each year ( $p < 0.05$ ).

The net DMY values (total DMY—WSC yield per plant) for the three C2 selected populations, C2<sup>S+</sup> 1.3±0.06 g, C2<sup>S-</sup> 0.725±0.049 g and C2<sup>S</sup> random 1.07±0.049 g, were found to be significantly different ( $p < 0.01$ ) indicating that the increase in biomass was not solely due to the increased WSC content.

Effect of WSC Selection on Total WSC Yield

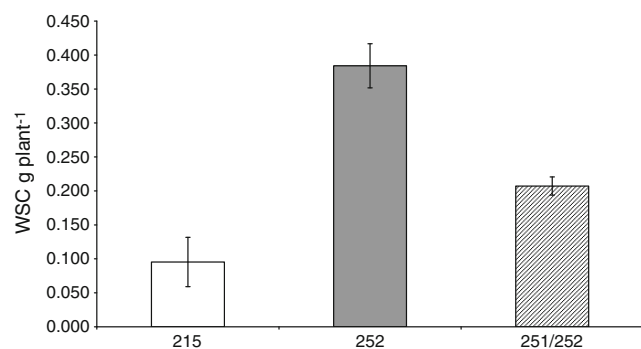
The C2<sup>S+</sup> plants that had undergone three rounds of high WSC selection exhibited a WSC yield value (WSC×DMY) of 3.9-fold higher than the C2<sup>S</sup> random plants that had not undergone WSC directed selection. By contrast the WSC yield for C2<sup>S-</sup> plants that had undergone three rounds of low WSC selection had decreased to 0.14-fold compared with the randomly selected plants (Table 2).

Effect of WSC Selection on Haplotype Frequency

Pairwise comparison of the haplotype frequencies in the three final C2<sup>S</sup> populations for each gene derived marker locus revealed that there were significant haplotype frequency differences for *LpsaINV1:4* (Table 2). F-statistic and exact test values of  $p < 0.0001$  and  $p < 0.0005$  were calculated for

*LpsaINV1:4* revealing significant pair wise differences between the C2<sup>S+</sup> and C2<sup>S-</sup> populations and the C2<sup>S+</sup> and C2<sup>S</sup> random populations. However, the C2<sup>S-</sup> low WSC and control C2<sup>S</sup> random populations were not found to differ significantly in haplotype frequency (F-statistic  $p > 0.25$ ; exact test  $p > 0.25$ ). The WSC yields of the three *LpsaINV1:4* haplotypes across the C2<sup>S</sup> populations as a whole (Fig. 2), demonstrated that plants with the *LpsaINV1:4/252* haplotype contained the highest WSC content; plants with the *LpsaINV1:4/251* haplotype had the lowest WSC, and *LpsaINV1:4/251/252* heterozygotic plants were intermediate for WSC content. A non glycoside hydrolase marker locus was also included in the study; cinnamate-4-hydroxylase (C4H) is the first enzyme in the lignin biosynthesis pathway, which can be considered an alternative sink for photosynthetic carbon. Transgenic plants with decreased lignin have been demonstrated previously to accumulate cellulose [12]. The respective F-statistic and exact test values of  $p < 0.0002$  and  $p < 0.00003$  for *LpC4H* haplotypes identified differences between the C2<sup>S+</sup> high and C2<sup>S-</sup> low WSC populations, but not between the C2<sup>S+</sup> and C2<sup>S</sup> random (F-statistic  $p > 0.6$ ; exact test,  $p > 0.3$ ) or the C2<sup>S</sup> random and the C2<sup>S-</sup> (F-statistic,  $p > 0.13$ ; exact test,  $p > 0.002$ ). The *LpC4H* haplotypes had therefore not undergone significant divergence





**Fig. 2** Total WSC yield of plants based on selection of *L. perenne* soluble-acid invertase 1:4 haplotypes 251 ( $n=43$ ), 252 ( $n=15$ ), and 251/252 ( $n=31$ ) distributed among the  $C2^S$  synthetic populations. Vertical bars represent the standard error of the mean

from the random control following three rounds of WSC directed selection.

*LpcwINV2* haplotype distributions were found to be significantly different between the control population  $C2^S$  random and  $C2^{S+}$  high WSC (F-statistic,  $p<0.00001$ ; exact test,  $p<0.00003$ ). However, pair wise differences were not observed between either the  $C2^S$  low WSC and  $C2^{S+}$  high WSC populations (F-statistic,  $p>0.01$ ; exact test,  $p>0.02$ ) or  $C2^S$  and the control population (F-statistic,  $p>0.03$ ; exact test,  $p>0.03$ ), indicating that *LpcwINV2* haplotypes had not significantly diverged. Differences in haplotype frequency distributions for *Lp6G-FFT1*, *LpcwINV1* and *LpsaINV5* were not identified between the three populations. The full molecular marker data set can be viewed under additional data online (see [Electronic supplementary material](#)).

The three WSC populations of *L. perenne* were statistically analysed by the  $\chi^2$  test for association between *LpsaINV1:4* haplotype and WSC level. The null hypothesis was formed that WSC directed selection had no effect on haplotype distribution. As such, *LpsaINV1:4* haplotypes would have been expected to be equally distributed in each population. However, the observed values significantly differed ( $\chi^2$   $p$  value  $<0.001$ ) from the expected and resulted in rejection of the null hypothesis. The *LpsaINV1:4/252* haplotype was absent from the  $C2^S$  population while being present in 14/30  $C2^{S+}$  high WSC plants and 1/30 control  $C2^S$  random plant (Table 2). The number of observed *LpsaINV1:4/251* SNP markers was found to significantly differ from the expected values for each population ( $\chi^2$   $p$  value  $<0.05$ ). Haplotype *LpsaINV1:4/251* was present in 21/30 low selected plants, 6/30 high selected plants and 16/30 randomly selected plants (Table 2), which resulted in the rejection of the null hypothesis for haplotype *LpsaINV1:4/251*. The proportion of heterozygous plants (haplotype *LpsaINV1:4/251/252*) was not significantly different between the three populations and was therefore in agreement with the null hypothesis.

## Discussion

### Increasing WSC Yield Through WSC Directed Selection

Each *L. perenne* breeding cycle requires between 3 to 5 years for completion to obtain and fix the desired phenotype [1]. Phenotypic selection based on WSC yield is a relatively easy trait to identify and quantify, enabling populations of *L. perenne* with high sugar yields to be generated with conventional breeding methods. However, the underlying genetic regulation of WSC accumulation in *Lolium* is complex, with QTL being present on chromosomes 1, 2, 5 and 6 [2, 3].

Following recombination and WSC directed selection distinct populations of plants with low and high WSC values on a dry weight basis were produced (Table 2). These data indicate that under WSC directed selection, *L. perenne* populations containing either high or low WSC content had undergone genetic divergence and were distinct from the WSC phenotypes present within the randomly selected population. The plants in this study were grown in a common environment and analysed in their first year of growth, it was not therefore possible to directly compare the  $C0$  population with the  $C2$  random population to analyse any random drift that had occurred over the 2 years. The effects of drift were confounded with possible environmental variation across selection cycles, but significant divergence between selected and random populations indicated that these effects were minor compared to selection responses.

A secondary effect of selecting high WSC plants was the observation of an increase in biomass with respect to the controls. The data in Table 2 demonstrate that following three cycles of recurrent WSC directed selection of a synthetic population, the total WSC yield in the  $C2^{S+}$  plants was 3.9-fold higher than the  $C2^S$  random plants. The underlying genetic complexity of WSC accumulation [2, 3, 13] appears to be associated with DMV. However, the alterations in WSC accumulation could not, in isolation, account for the concomitant divergence in the DMV values observed, in this study. The identification and combination of major QTL associated with early spring DMV and WSC would be of value to the production of bio-ethanol production from perennial ryegrass.

### Associating Genotype with Phenotype and Generating Molecular Markers for High WSC Selection

A significant difference in one gene locus was identified in phenotypes with both high biomass per plant and elevated WSC yields. Of the six candidate genes identified, the haplotype frequencies of *Lp6G-FFT1*, *LpcwINV1*, *LpsaINV5*, *LpcwINV2* and *LpC4H* genes were not found to have

undergone significant divergence as a result of WSC directed selection, consistent with Sköt et al. [14] who demonstrated no association of the *LpcAI* gene with WSC content. However, haplotypes of the *LpsaINV1:4* gene had undergone divergence following WSC directed selection. The lowest frequency of haplotype *LpsaINV1:4/251* was exhibited by the  $C2^{S+}$  populations, suggesting that, although significant differences between the three populations were observed (Table 2), this marker was not suitable for the identification and selection of high WSC phenotypes. However, glycoside hydrolase *LpsaINV1:4/252*, which is a soluble vacuolar invertase, was predominantly associated with the high WSC population ( $p < 0.001$ ) (Table 2). The observed numbers of the heterozygous *LpsaINV1:4/251/252* haplotype did not statistically differ from the expected numbers in each of the three populations. Only when considered in isolation was the SNP *LpsaINV1:4/252* found to be significantly associated with high WSC levels. Furthermore, this shift in frequency of the 252 allele was unlikely to be due to drift, as it is apparent that, the large increase in the frequency of the 252 allele in the high WSC population distinguishes it from both the low WSC and random populations (Fig. 3). It was apparent that following recombination and selection, divergence between the populations had occurred and *LpsaINV1:4/252* segregated in the  $C2^{S+}$  population.

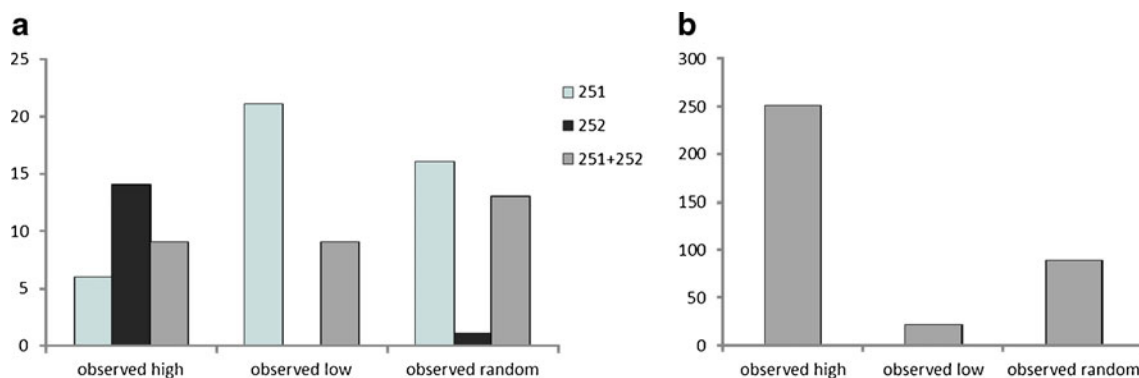
Expression of acid *LpsaINV1:4* mRNA has been shown previously to be mainly located in leaf sheaths, with comparatively minor expression in leaf blades [15], concurrent with WSC accumulation. The majority of plant tissue excised for WSC quantification in this study was leaf blade, although some sheaths were present. While SNP *LpsaINV1:4/252* was associated with high WSC yields (Table 2), it is not known whether this SNP was present as a functional marker, potentially acting to effect sugar translocation, or whether it is a non-functional linked marker.

The development of molecular markers for high WSC content and high biomass increases the potential of this abundant temperate grass species to produce a sustainable

source of renewable liquid fuel. Chromosome 6 in the *L. perenne* genome is the location of both a major QTL for autumn leaf blade WSC where SSR marker *rv1423* underlies the QTL [2] and a separate one for spring leaf blade WSC where *RZ28* and cytoplasmic *Alk Inv1/4* underlie the QTL [2]. Only spring leaf WSC was targeted in this study but the recombination frequency between the two was in the order of 2–4% in another study [3]. Turner et al. [16] assessed changes in a limited number of SSR allele frequencies during the current selection experiment. Three markers on chromosome 6 (*rv1423*, *rv0641* and *rv0739*) were analysed by Genepop routines and all demonstrated a *P* value of 0.00001 showing significant allelic divergence during selection at loci spanning the whole of this chromosome. Haplotype *LpsaINV1:4/252* may have been inherited with the selected alleles of *rv1423*, *rv0641* and *rv0739* [16], which were mostly the high sugar-associated alleles from LTS18. Further mapping and analysis of this chromosomal region should reveal the markers functionally associated with the observed high sugar yield QTL.

#### Sugar Yield and Conversion to Bio-ethanol

Elite germplasm can be produced from out-breeding perennial grasses as they are amenable to genetic improvement through phenotypic recurrent selection [1]. More recently, Gallagher et al. [4] highlighted that grasses containing high WSC levels could have the potential to fulfil a niche in the biofuel industry. Of considerable long term importance in the bio-generation of liquid fuel will be not only the total ethanol yield (energy out), but also the ease with which the available plant sugars can be extracted for conversion into ethanol (energy in; [17]). Annual dry matter yields of *L. perenne* have reached  $23.1 \text{ t ha}^{-1}$  containing average WSC levels of 33.4% [1] equating to  $7.7 \text{ t ha}^{-1}$  of readily extractable WSC. Based on the theoretical maximum yield of ethanol per kg of sucrose being 0.51 kg ethanol plus 0.49 kg  $\text{CO}_2$  [18] and assuming complete microbial conversion of sugars to



**Fig. 3** a *LpsaINV1:4* allele frequency and b WSC content (mg g<sup>-1</sup> DM) in the selected high, low, and random  $C2^{S+}$  populations ( $n=30$ )

alcohol, one hectare of grassland has the potential to produce 3.9 tonnes of ethanol, or 170.34 kg of ethanol per tonne of grass through conversion of the WSC alone, which requires relatively little energy input especially if grown as a mixed sward with a forage legume such as white clover. In practice, ethanol streams in excess of 4% have been produced from the WSC contained in fresh grass juice following the enzymatic liberation of fructose and glucose using fructan hydrolase of *Lactobacillus plantarum* [19].

Concurrent developments are being made in the bioconversion process which will increase the efficiency; for example, Martel et al. [20] reported on the over expression of a *Lactobacillus paracasei*  $\beta$ -fructosidase: an enzyme capable of hydrolysing polymeric fructan, the major WSC component of *L. perenne*, into monomeric sugars, available for fermentation, and Martel et al. [21] describe the expression of bacterial levanase in yeast which enables simultaneous saccharification and fermentation of grass juice to ethanol. Based on the above yield and conversion data [1, 11, 22, 23], one tonne of perennial ryegrass could potentially produce over 280 kg of ethanol.

Following WSC removal, the lignocellulosic fraction of *L. perenne* has been reported to consist of 32.8% to 36.8% cellulose and 41.1% hemicellulose on a dry weight basis [22, 23]. Subsequent lignocellulosic digestion of the biomass residue remaining following juicing would increase the total sugar yield and thereby elevate the potential yield of bio-ethanol per tonne. The identification of *L. perenne* plants with high biomass and elevated WSC and the elucidation of the genetic regulation underlying these traits is of great significance for increasing the sustainable production of bio-ethanol from perennial ryegrass.

While a biotechnological case exists for the use of transgenics for increasing biofuel production, in terms of plant material and the use of recombinant enzymes and micro-organisms during processing [24], the exploitation of natural genetic variation also has significant potential. For example, the identification of a haplotype which is associated with elevated WSC in *L. perenne* offers the potential to optimise and introgress the high WSC trait into varieties targeted at bioenergy or dual use applications.

**Acknowledgements** The work reported in this study was funded by the European Union Framework IV project GRASP and the UK Biotechnology and Biological Research Council (BBSRC) grant number SPG 03134 CFC.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

1. Wilkins P, Humphreys MO (2003) Progress in breeding perennial forage grasses for temperate agriculture. *J Agric Sci* 140(2):129–150
2. Turner LB, Cairns AJ, Armstead IP, Ashton J, Skøt K, Whittaker D et al (2006) Dissecting the regulation of fructan metabolism in perennial ryegrass (*Lolium perenne*) with quantitative trait locus mapping. *New Phytol* 169(1):45–58
3. Turner LB, Farrell M, Humphreys MO, Dolstra O (2010) Testing water-soluble carbohydrate QTL effects in perennial ryegrass (*Lolium perenne* L.) by marker selection. *Theor Appl Genet* 121(8):1405–17
4. Gallagher JA, Cairns AJ, Turner LB (2007) Fructan in temperate forage grasses; agronomy, physiology and molecular biology. In: *Recent advances in fructooligosaccharides research*. Research Signpost, Kerala. pp 15–46. ISBN: 81-308-0146-9
5. Lopez-Casado G, Urbanowicz BR, Damasceno CM, Rose JK (2008) Plant glycosyl hydrolases and biofuels: a natural marriage. *Curr Opin Plant Biol* 11(3):329–337
6. Posselt UK, Barre P, Brazauskas G, Turner LB (2006) Comparative analysis of genetic similarity between perennial ryegrass genotypes investigated with AFLPs, ISSRs, RAPDs and SSRs. *Czech J Genet Plant Breed* 42(3):87–94
7. Turner LB, Humphreys MO, Cairns AJ, Pollock CJ (2002) Carbon assimilation and partitioning into non-structural carbohydrate in contrasting varieties of *Lolium perenne*. *J Plant Physiol* 159:257–263
8. Laurentin A, Edwards CA (2003) A microtiter modification of the anthrone-sulfuric acid colorimetric assay for glucose-based carbohydrates. *Anal Biochem* 315:143–145
9. Farrar K, Asp T, Lübberstedt T, Xu M, Thomas A, Christiansen C et al (2007) Construction of two *Lolium perenne* BAC libraries and identification of BACs containing candidate genes for disease resistance and forage quality. *Mol Breed* 19(1):15–26
10. Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47–50
11. Chernoff H, Lehmann EL (1954) The use of maximum likelihood estimates in  $\chi^2$  tests for goodness-of-fit. *Ann Math Stat* 25:579–586
12. Hu WJ, Harding SA, Lung J, Popko JL, Ralph J, Stokke DD et al (1999) Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. *Nat Biotechnol* 17(8):808–12
13. Turner LB, Cairns AJ, Armstead IP, Thomas H, Humphreys MW, Humphreys MO (2008) Does fructan have a functional role in physiological traits? Investigation by quantitative trait locus mapping. *New Phytol* 179(3):765–775
14. Skøt L, Humphreys J, Humphreys MO, Thorogood D, Gallagher J, Sanderson R et al (2007) Association of candidate genes with flowering time and water-soluble carbohydrate content in *Lolium perenne* (L.). *Genetics* 177:535–547
15. Gallagher JA, Cairns AJ, Pollock CJ (2004) Cloning and characterization of a putative fructosyltransferase and two putative invertase genes from the temperate grass *Lolium temulentum* L. *J Exp Bot* 55:557–569
16. Turner LB, Skøt KP, Farrell M, Cairns AJ, Armstead IP, Farrar K et al. (2009) Changes in SSR allele frequencies during phenotypic selection for water-soluble carbohydrate (WSC) in an experimental population of perennial ryegrass. In: Lübberstedt, T. et al. (ed) *Proceedings, 27th EUCARPIA Symposium on Improvement of Fodder Crops and Amenity Grasses*, Copenhagen, 19–23 August 2007. p. 27
17. Bouton JH (2007) Molecular breeding of switchgrass for use as a biofuel crop. *Curr Opin Genet Dev* 17:553–558



18. Hamelinck CN, van Hooijdonk G, Faaij APC (2005) Ethanol from lignocellulosic biomass: techno-economic performance in short-, middle- and long-term. *Biomass & Bioenergy* 28(4):384–410
19. Morris SM, Jackson CJ, Gallagher JA, Kelly SL, Donnison IS (2008) High Sugar Perennial Rye Grass as a Bioethanol Feedstock \$ An Alternative Use for Forage Grasses. In: Proceedings of the 16th European Biomass Conference and Exhibition, Valencia, Spain, 2–6th June 2008. ETA-Florence Renewable Energies, Florence, Italy, pp. 60–66
20. Martel CM, Warrilow AGS, Jackson CJ, Mullins JGL, Togawa RC, Parker JE et al (2010) Expression, purification and use of the soluble domain of *Lactobacillus paracasei*  $\beta$ -fructosidase to optimise production of bioethanol from grass fructans. *Bioresour Technol* 101:4395–4402
21. Martel CM, Parker JE, Jackson CJ, Warrilow AGS, Rolley N, Greig C et al (2011) Expression of bacterial levanase in yeast enables simultaneous saccharification and fermentation of grass juice to bioethanol. *Bioresour Technol* 102:1503–1508
22. Liu CF, Xu F, Sun JX, Ren JL, Curling S, Sun RC et al (2006) Physicochemical characterization of cellulose from perennial ryegrass leaves (*Lolium perenne*). *Carbohydr Res* 341:2677–2687
23. Xu F, Geng ZC, Sun JX, Liu CF, Ren JL, Sun RC et al (2006) Fractional and structural characterization of hemicelluloses from perennial ryegrass (*Lolium perenne*) and cocksfoot grass (*Dactylis glomerata*). *Carbohydr Res* 341:2073–2082
24. Gressel J (2008) Transgenics are imperative for biofuel crops. *Plant Sci* 174:246–263